(*e*) translation and rotation functions (see Chapter 2.3), when defined in direct space, always have their counterpart in reciprocal space.

2.2.7. Scheme of procedure for phase determination

A traditional procedure for phase assignment may be schematically presented as follows:

Stage 1: *Normalization of s.f.'s.* See Section 2.2.4.

Stage 2: *(Possible) estimation of one-phase s.s.'s.* The computing program recognizes the one-phase s.s.'s and applies the proper formulae (see Section 2.2.5.9).

Each phase is associated with a reliability value, to allow the user to regard as known only those phases with reliability higher than a given threshold.

Stage 3: *Search of the triplets.* The reflections are listed for decreasing $|E|$ values and, related to each $|E|$ value, all possible triplets are reported (this is the so-called $\sum_2$ list). The value $G = 2|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h-k}}|/\sqrt{N}$ is associated with every triplet for an evaluation of its efficiency. Usually reflections with $|E| < E_s$ (E_s may range from 1.2 to 1.6) are omitted from this stage onward.

Stage 4: *Definition of the origin and enantiomorph.* This stage is carried out according to the theory developed in Section 2.2.3. Phases chosen for defining the origin and enantiomorph, one-phase seminvariants estimated at stage 2, and symbolic phases described at stage 5 are the only phases known at the beginning of the phasing procedure. This set of phases is conventionally referred to as the *starting set*, from which iterative application of the tangent formula will derive new phase estimates.

Stage 5: *Assignment of one or more (symbolic or numerical) phases.* In complex structures the number of phases assigned for fixing the origin and the enantiomorph may be inadequate as a basis for further phase determination. Furthermore, only a few one-phase s.s.'s can be determined with sufficient reliability to make them qualify as members of the starting set. Symbolic phases may then be associated with some (generally from 1 to 6) high-modulus reflections (symbolic addition procedures). Iterative application of triplet relations leads to the determination of other phases which, in part, will remain expressed by symbols (Karle & Karle, 1966).

In other procedures (multisolution procedures) each symbol is assigned four phase values in turn: $\pi/4, 3\pi/4, 5\pi/4, 7\pi/4$. If p symbols are used, in at least one of the possible 4^p solutions each symbolic phase has unit probability of being within 45° of its true value, with a mean error of 22.5°.

To find a good starting set a *convergence method* (Germain *et al.*, 1970) is used according to which: (*a*)

$$\langle\alpha_{\mathbf{h}}\rangle = \sum_j G_j I_1(G_j)/I_0(G_j)$$

is calculated for all reflections (j runs over the set of triplets containing **h**); (*b*) the reflection is found with smallest $\langle\alpha\rangle$ not already in the starting set; it is retained to define the origin if the origin cannot be defined without it; (*c*) the reflection is eliminated if it is not used for origin definition. Its $\langle\alpha\rangle$ is recorded and $\langle\alpha\rangle$ values for other reflections are updated; (*d*) the cycle is repeated from (*b*) until all reflections are eliminated; (*e*) the reflections with the smallest $\langle\alpha\rangle$ at the time of elimination go into the starting set; (*f*) the cycle from (*a*) is repeated until all reflections have been chosen.

Stage 6: *Application of tangent formula.* Phases are determined in reverse order of elimination in the convergence procedure. In order to ensure that poorly determined phases $\varphi_{\mathbf{k}_j}$ and $\varphi_{\mathbf{h-k}_j}$ have little effect in the determination of other phases a weighted tangent formula is normally used (Germain *et al.*, 1971):

$$\tan\varphi_{\mathbf{h}} = \frac{\sum_j w_{\mathbf{k}_j} w_{\mathbf{h-k}_j} |E_{\mathbf{k}_j} E_{\mathbf{h-k}_j}| \sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h-k}_j})}{\sum_j w_{\mathbf{k}_j} w_{\mathbf{h-k}_j} |E_{\mathbf{k}_j} E_{\mathbf{h-k}_j}| \cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h-k}_j})}, \qquad (2.2.7.1)$$

where

$$w_{\mathbf{h}} = \min\,(0.2\alpha, 1).$$

Once a large number of contributions are available in (2.2.7.1) for a given $\varphi_{\mathbf{h}}$, then the value of $\alpha_{\mathbf{h}}$ quickly becomes greater than 5, and so assigns an unrealistic unitary weight to $\varphi_{\mathbf{h}}$. In this respect a different weighting scheme may be proposed (Hull & Irwin, 1978) according to which

$$w = \psi \exp(-x^2) \int_0^x \exp(t^2)\,\mathrm{d}t, \qquad (2.2.7.2)$$

where $x = \alpha/\langle\alpha\rangle$ and $\psi = 1.8585$ is a constant chosen so that $w = 1$ when $x = 1$. Except for ψ, the right-hand side of (2.2.7.2) is the Dawson integral which assumes its maximum value at $x = 1$ (see Fig. 2.2.7.1): when $\alpha > \langle\alpha\rangle$ or $\alpha < \langle\alpha\rangle$ then $w < 1$ and so the agreement between α and $\langle\alpha\rangle$ is promoted.

Alternative weighting schemes for the tangent formula are frequently used [for example, see Debaerdemaeker *et al.* (1985)]. In one (Giacovazzo, 1979*b*), the values $\alpha_{\mathbf{k}_j}$ and $\alpha_{\mathbf{h-k}_j}$ (which are usually available in direct procedures) are considered as additional *a priori* information so that (2.2.7.1) may be replaced by

$$\tan\varphi_{\mathbf{h}} \simeq \frac{\sum_j \beta_j \sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h-k}_j})}{\sum_j \beta_j \cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h-k}_j})}, \qquad (2.2.7.3)$$

where β_j is the solution of the equation

$$D_1(\beta_j) = D_1(G_j) D_1(\alpha_{\mathbf{k}_j}) D_1(\alpha_{\mathbf{h-k}_j}). \qquad (2.2.7.4)$$

In (2.2.7.4),

$$G_j = 2|E_{\mathbf{h}} E_{\mathbf{k}_j} E_{\mathbf{h-k}_j}|\sqrt{N}$$

or the corresponding second representation parameter, and $D_1(x) = I_1(x)/I_0(x)$ is the ratio of two modified Bessel functions.

In order to promote (in accordance with the aims of Hull and Irwin) the agreement between α and $\langle\alpha\rangle$, the distribution of α may be used (Cascarano, Giacovazzo, Burla *et al.*, 1984; Burla *et al.*, 1987); in particular, the first two moments of the distribution: accordingly,

$$w = \left\{ \exp\left[\frac{-(\alpha - \langle\alpha\rangle)^2}{2\sigma_\alpha^2} \right] \right\}^{1/3}$$

may be used, where σ_α^2 is the estimated variance of α.

Stage 7: *Figures of merit.* The correct solution is found among several by means of figures of merit (FOMs) which are expected to be extreme for the correct solution. Largely used are (Germain *et al.*, 1970)

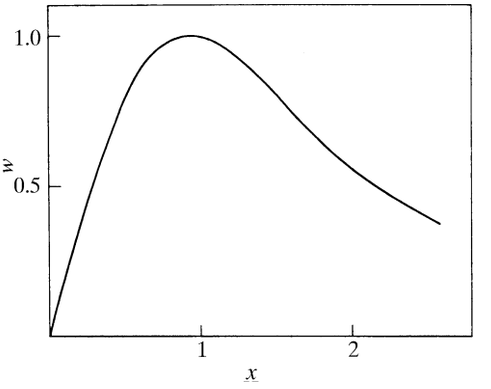

Fig. 2.2.7.1. The form of w as given by (2.2.7.2).

(*a*) $$\text{ABSFOM} = \sum_{\mathbf{h}} \alpha_{\mathbf{h}} / \sum_{\mathbf{h}} \langle \alpha_{\mathbf{h}} \rangle,$$

which is expected to be unity for the correct solution.

(*b*) $$\text{PSI0} = \frac{\sum_{\mathbf{h}} \left| \sum_{\mathbf{k}} E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}} \right|}{\sum_{\mathbf{h}} \left(\sum_{\mathbf{k}} |E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}|^2 \right)^{1/2}}.$$

The summation over **k** includes (Cochran & Douglas, 1957) the strong $|E|$'s for which phases have been determined, and indices **h** correspond to very small $|E_{\mathbf{h}}|$. Minimal values of PSI0 (≤ 1.20) are expected to be associated with the correct solution.

(*c*) $$R_\alpha = \frac{\sum_{\mathbf{h}} |\alpha_{\mathbf{h}} - \langle \alpha_{\mathbf{h}} \rangle|}{\sum_{\mathbf{h}} \langle \alpha_{\mathbf{h}} \rangle}.$$

That is, the Karle & Karle (1966) residual between the actual and the estimated α's. After scaling of $\alpha_{\mathbf{h}}$ on $\langle \alpha_{\mathbf{h}} \rangle$ the correct solution should be characterized by the smallest R_α values.

(*d*) $$\text{NQEST} = \sum_j G_j \cos \Phi_j,$$

where G is defined by (2.2.5.21) and

$$\Phi = \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{l}} - \varphi_{\mathbf{h}-\mathbf{k}-\mathbf{l}}$$

are quartet invariants characterized by large basis magnitudes and small cross magnitudes (De Titta *et al.*, 1975; Giacovazzo, 1976). Since G is expected to be negative as well as $\cos \Phi$, the value of NQEST is expected to be positive and a maximum for the correct solution.

Figures of merit are then combined as

$$\begin{aligned} \text{CFOM} = w_1 & \frac{\text{ABSFOM} - \text{ABSFOM}_{\min}}{\text{ABSFOM}_{\max} - \text{ABSFOM}_{\min}} \\ & + w_2 \frac{\text{PSI0}_{\max} - \text{PSI0}}{\text{PSI0}_{\max} - \text{PSI0}_{\min}} \\ & + w_3 \frac{R_{\alpha_{\max}} - R_\alpha}{R_{\alpha_{\max}} - R_{\alpha_{\min}}} \\ & + w_4 \frac{\text{NQEST} - \text{NQEST}_{\min}}{\text{NQEST}_{\max} - \text{NQEST}_{\min}}, \end{aligned}$$

where w_i are empirical weights proportional to the confidence of the user in the various FOMs.

Different FOMs are often used by some authors in combination with those described above: for example, enantiomorph triplets and quartets are supplementary FOMs (Van der Putten & Schenk, 1977; Cascarano, Giacovazzo & Viterbo, 1987).

Different schemes of calculating and combining FOMs are also used: a recent scheme (Cascarano, Giacovazzo & Viterbo, 1987) uses

(*a*1) $$\text{CPHASE} = \frac{\sum w_j G_j \cos(\Phi_j - \theta_j) + w_j G_j \cos \Phi_j}{\sum_{\text{s.i.+s.s.}} w_j G_j D_1(G_j)},$$

where the first summation in the numerator extends over symmetry-restricted one-phase and two-phase s.s.'s (see Sections 2.2.5.9 and 2.2.5.10), and the second summation in the numerator extends over negative triplets estimated *via* the second representation formula [equation (2.2.5.13)] and over negative quartets. The value of CPHASE is expected to be close to unity for the correct solution.

(*a*2) $\alpha_{\mathbf{h}}$ for strong triplets and $E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}$ contributions for PSI0 triplets may be considered random variables: the agreements between their actual and their expected distributions are considered as criteria for identifying the correct solution.

(*a*3) correlation among some FOMs is taken into account.

According to this scheme, each FOM (as well as the CFOM) is expected to be unity for the correct solution. Thus one or more figures are available which constitute a sort of criterion (on an absolute scale) concerning the correctness of the various solutions: FOMs (and CFOM) $\simeq 1$ probably denote correct solutions, CFOMs $\ll 1$ should indicate incorrect solutions.

Stage 8: *Interpretation of E maps.* This is carried out in up to four stages (Koch, 1974; Main & Hull, 1978; Declercq *et al.*, 1973):

(*a*) peak search;

(*b*) separation of peaks into potentially bonded clusters;

(*c*) application of stereochemical criteria to identify possible molecular fragments;

(*d*) comparison of the fragments with the expected molecular structure.

2.2.8. Other multisolution methods applied to small molecules

In very complex structures a large initial set of known phases seems to be a basic requirement for a structure to be determined. This aim can be achieved, for example, by introducing a large number of permutable phases into the initial set. However, the introduction of every new symbol implies a fourfold increase in computing time, which, even in fast computers, quickly leads to computing-time limitations. On the other hand, a relatively large starting set is not in itself enough to ensure a successful structure determination. This is the case, for example, when the triplet invariants used in the initial steps differ significantly from zero. New strategies have therefore been devised to solve more complex structures.

(1) *Magic-integer methods*

In the classical procedure described in Section 2.2.7, the unknown phases in the starting set are assigned all combinations of the values $\pm\pi/4$, $\pm 3\pi/4$. For n unknown phases in the starting set, 4^n sets of phases arise by quadrant permutation; this is a number that increases very rapidly with n. According to White & Woolfson (1975), phases can be represented for a sequence of n integers by the equations

$$\varphi_i = m_i x \pmod{2\pi}, \quad i = 1, \ldots, n. \tag{2.2.8.1}$$

The set of equations can be regarded as the parametric equation of a straight line in n-dimensional phase space. The nature and size of errors connected with magic-integer representations have been investigated by Main (1977) who also gave a recipe for deriving magic-integer sequences which minimize the r.m.s. errors in the represented phases (see Table 2.2.8.1). To assign a phase value, the variable x in equation (2.2.8.1) is given a series of values at equal intervals in the range $0 < x < 2\pi$. The enantiomorph is defined by exploring only the appropriate half of the n-dimensional space.

A different way of using the magic-integer method (Declercq *et al.*, 1975) is the *primary–secondary P–S method* which may be described schematically in the following way:

(*a*) Origin- and enantiomorph-fixing phases are chosen and some one-phase s.s.'s are estimated.

(*b*) Nine phases [this is only an example: very long magic-integer sequences may be used to represent primary phases (Hull *et al.*, 1981; Debaerdemaeker & Woolfson, 1983)] are represented with the approximated relationships:

references